Drug component analysis based on Fuzzified Decision tree optimized neural network for disease related drug recommendation

E.Loganathan[1], M. Prakash[2*], G.Sivakumar[3]
[1]*Assistant Professor, Department of Computer Engineering, Erode Sengunthar Engineering College (Autonomous) Perundurai, Erode – 638 057*
[2]*Master of Engineering in Computer Science and Engineering, Erode Sengunthar Engineering College (Autonomous) Perundurai, Erode – 638 057*
[3]*Head of the Department, Department of Computer Science and Engineering, Erode Sengunthar EngineeringCollege (Autonomous), Perundurai, Erode-638057*

***Abstract*—Drug component analysis is an important factor for recommending disease oriented treatment for handling patients. Due to improper combination drugs prescription in medical industries cause more side effect by taking the medicinal drugs. So the combination of related feature analysis is main thing for identifying the correct the combinational facts. Most of the previous methodologies doesn't carry to find drug relation to make combination. To resolve this problem, we propose a Drug feature penetration similarity rate (DFPSR)and deep Fuzzified Decision tree optimized neural network (FDTONN) classifier is applied to find the relational drug combination factor for recommending correct drug to the patients. The selected features are trained with FDTONN. This selected the drug pattern relation based on disease prone weight recommending the best features. The proposed system achieves high performance compared to the other existing system to solving the feature dimensionality ratio to make higher precision, recall rate, f-measure in redundant time complexity. This produce best performance as well other methods to attain best recommendation to the patients.**

***Keywords*— drug data analysis: ANN, Fuzzy, feature selection and classification, machine learning.**

I. INTRODUCTION

Health data and drug data process in medicinal organizational unit and is accessed through modern information solutions to provide additional insights to executives for finding relational drugs. For example, several groups have performed disease prediction and drug analysis using data residing on different data servers located in different geographic locations. However, IT solutions allow access to data from other locations to generate intelligence.

The data dimension becomes another form of so-called big data leads feature dimensionality analysis, which is massive but in architecture is so tedious for finding the relational drug analysis. Medical Organizations maintain such vast amounts of data at different locations on any network that they can access to generate intelligence for decision-making or anything else. To provide access to big data, several previous methods have been previously discussed by various researchers. However, these methods still have many problems. Healthcare organizations maintain disparate big data at remote locations that are used to perform analytics to generate results that support decision making. Such data was accessed through new network communications.

In this paper the contribution is to develop an efficient Drug component analysis based on Fuzzified Decision tree optimized neural network for disease related drug recommendation. Drug feature penetration similarity rate (DFPSR)and deep Fuzzified Decision tree optimized neural network (FDTONN) classifier is applied to find the relational drug combination factor for recommending correct drug to the patients. The selected features are trained with FDTONN. This selected the drug pattern relation based on disease prone weight recommending the best features. The proposed system achieves high performance compared to the other existing system to solving the feature dimensionality ratio to make higher precision, recall rate, f-measure in redundant time complexity.

II. RELATED WORK

Machine Learning Algorithms and Techniques Machine learning techniques are used in various medical databases to automatically explore large and complex data for cardiac prognosis. Recently, many researchers have been using machine learning techniques to help medical professionals and specialists in diagnosing heart-related diseases. Most clustering-based decision frameworks are easy to understand supervised learning methods for neural classifiers. It handles numeric and attribute datasets. The internal nodes of each branch consist of branches and leaf nodes representing the eigenvalues of the given dataset and decision tree, i.e. a tree structure with an experimental internal terminal. View classes that see leaf projections or effects show nodes. Classification rules start from node nodes based on predictive properties and some constraints.

A Machine Learning-Based Disease Prediction Algorithm Using Synthetic Datasets. Approaches to detecting drug-related components in machine learning using ensembles for disease prediction. Similarly, machine learning provides a hybrid approach to predicting heart disease. This method combines various techniques for disease prediction. Using PSO-based clustering optimization gives higher clustering accuracy. Density-based clustering techniques support symptom-based disease prediction. A disease prognostic scheme based on machine learning algorithms using Rough K-Means (RKM) clustering is introduced to detect chronic diseases. The system learning machine proposed in this paper describes the IoT form of a neural network algorithm for predicting heart disease. Perceptron (MLP) training data collection experiment for IoT. In this approach, both inputs and outputs are masked layer by layer for a production output and one or more layers, where there are multiple layers of such inputs.

III. PROPOSED METHODOLOGY

Towards the development of drug pattern and relation feature is analyzed based on ANFIS model optimized with feature selection and classification model. propose a Drug feature penetration similarity rate (DFPSR)and deep Fuzzified Decision tree optimized neural network (FDTONN) classifier is applied to find the relational drug combination factor for recommending correct drug to the patients.

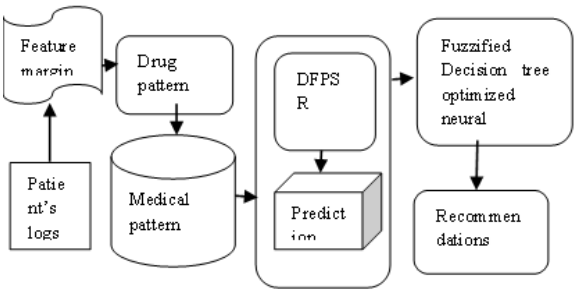

Figure 1: Proposed architecture diagram FDTONN

The selected features are trained with FDTONN. This selected the drug pattern relation based on disease prone weight recommending the best features. Fig. 1 shows the proposed architecture diagram FDTONN. The proposed system achieves high performance compared to the other existing system to solving the feature dimensionality ratio to make higher precision, recall rate, f-measure in redundant time complexity. This produce best performance as well other methods to attain best recommendation to the patients.

A. Preprocessing dataset

The drug compound dataset is collected from medical industry which contains the features related to medicinal chemical compounds and its composition details based on the disease combination relation. These dataset is filled by records with labels. Preprocessing verify the label presence of the feature presence and threshold margins. This normalize the dataset for analyzing the values to make perspective decision approach.

Algorithm

Input: Initialize Drug compounds dataset- Cds

Output: Normalized Drug dataset -Nds

Step 1: for all check the Labels Cds → cds1, cds2,…

Step 2: initialize I=0 for all margins drug feature check if == margin

Step 3: Compute to verify the range of feature components value.

Check ideal and actual margin range

If true==Cds

Create feature Fds→$f(x) = \sum_{n=1}^{\infty}(Actual\ margin)$

End if

Step 4: Nds ← normalized records

Step 5: end if end for;

The stage of preprocessing clarified in the above algorithm reduces the dimension based on the attributes by filtering the values. Each record contains attributes that represent information about an individual patient.

B. Drug feature penetration similarity rate (DFPSR)

Feature penetration similarity clustering group's large data sets under specific class names. As the dimensions and volume of data are enormous, it is necessary to group them into a specific category. A feature penetration similarity clustering method is proposed to perform drug pattern based on relation measurement. This method measures the similarity of features at multiple levels. From similarity, the procedure calculates feature complexity similarity (FDS) based on the number of models close to the feature and the total number of models in any class.

Correspondingly, the numeral of features is adjacent to what is accessible. Initially, a set of illustrations is allocated to each illness class collection, and the technique additional computes the FDS metric for grouping.

FDS Clustering Algorithm:

Input: Data Set Bds

Output: Cluster Set Cs

Start

Read Bds.

Initialize number of clusters Nc = ∑ Disease Classes

For each disease class n

Assign random samples.

CS(n) = ∑ Random Sample (Bds)

End

For each sample s

For each cluster c

For each cluster sample cls

Compute Feature Depth Similarity FDS = $\frac{\sum_{i=1}^{size(S)} S(i).value == Cls(i).value}{size(S)}$

End

Compute cumulative feature depth similarity CFDs.

CFDS = $\sum_{i=1}^{size(c)} C(i).FDS \Big/ size(C)$

End

Cluster c = Choose the cluster with maximum CFDs.

Index the data point to the selected cluster.

End

Stop

The methods discussed above indicate how big data is clustered based on feature similarity. Similarity is measured based on the depth of feature similarity available between different data points of any cluster. Based on the depth similarity values of the feature, a single set is selected and the data points are coded into the selected group.

C. Fuzzified Decision tree optimized neural network

Considering this, the feature weights will be adjusted so that the actual output matches the maximum measured weight based on the fuzzy optimized neural network. These results are repeated until the ANN returns the desired result. This optimization feature selects the most important and closest components to select input processing. The search formulation minimizes this overall error by matching standard features of neural networks with conventional weighted sum discrimination. The results are tuned to select the optimal weights using the proposed neural network genetic optimization technique.

Input: Drug pattern features

Output: scaling feature values

Step 1: Compute Fuzzy membership function on Rule set Rs

If check margin Cluster Ci

Max ideal margin > ideal values

Choose the Logic condition For feature patterns

End if

Step 2: Compute the Process of ANN Cross hidden layers

Step 3:choose the susbset feature index margins to create Fuzzy Logic Units in hidden layers as Ci←Rs

Step 4: Select the feature margins to the scalable units

Fsi(Fst→w.xi+b)>= y-€i, €y ≥ 0, 1 ≤ y ≤ n

Step 5: Compute actual drug pattern fsi as Max feature Mx

Step 6: Choose all the drug relation pattern Length

Step 7: Select all the best features F as cluster Fci

Fci→ min $\frac{1}{2}\left||w|\right| + c\sum_{i=1}^{n} €i$

Step 8: compute the ideal margin and relation factors xci,j
Step 9 Choose the max support drug relation features xnew,ci,j

Indexing class $Xnew, ci, j = Xci, j + \acute{\alpha} * (Xbest, ci - Xci, j) * r^2$

Step 10: select the margins if xci,j= x is the best,ci class then

Chioose best calss Xnew,ci = β*Ccenter,ci

Step 11: Complement lesson Maxci,j and apprise xnew,ci,j

$$X_{center,ci,d} = \frac{1}{nci} \sum_{l=1}^{nci} xci, l, d$$

Step 12: end if
Step 13: end for
Step 14: end for
Step 15: For all Drug pattern Class Drci→ group into cluster Ci
Step 16: Create threshold class based on best case fitness model min- max

$$X_{worst,ci}=X_{min}+(X_{max}-X_{min}+1).rand\ (Ci)$$

Step 17: End for
Step18: Compute the scaling margin mean weight split the class
Step 20: Split Max weight class by weight Cl→Max-W

Recurrent layers are optimized for connected hidden layers, such as NN layers that contain cross over fuzzy processing layers. Ideal processing values of drug relation weight is used to categorize Drug class recommendations by transaction. Refinement data all selected features are applied to the neural inputs and appear in the search for the best parent weight. Neural connectivity has specific targets in different modes of activity as directed by the driver. The target result is achieved through a classifier that classifies diseases based on recommendations.

D. Drug pattern recommendation (DPR)

The medical practitioner can recommend the treatment set to the patient according to the request. The value of the disease-prone factor (PDF) is measured according to the feature level similarity (FLS) measured on each sample available under the cluster. Based on that a set of recommendations are generated for the user.

Algorithm:

Input: Disease Class set Dcs, Cluster C, Test Sample s
Output: Recommendation R
Start
Read DCS, C
For each disease class dc
Compute disease-prone factor DPF.
PDF = $\sum_{i=1}^{Size(C(dc))} Features(C((S)DCI) \equiv Features(s) \Big/ size(C(dc))$
End
Choose class C with a full PDF.
For each treatment for class C
Compute success rate SR = $\sum_{i=1}^{size(C)} C(i).State == Success \Big/ size(C)$
End
Recommendation = Sort the treatment according to success rate.
Stop

The algorithm above measures the disease predisposition of a given test sample and the success rate of different treatment options to recommend to physicians.

IV. RESULT AND DISCUSSION

The proposed FDTONN drug component analysis and recommendation generation models are implemented and evaluated for their performance. The performance evaluation is carried in Python and the methods are evaluated on various metrics. The Drug Indications Database is collected from an online UCI repository. It contains 64 attributes: drug id, drug name, compound molecules (Chemical reaction) level, chemical id, and so on.

Table 1: Parameters settings

Parameters items	Values
Language	Python
Tool	Anaconda
Name of the dataset	Healthcare Drug Indications Database
Number of records	500000

The above table 1 shows the parameters settings for implantation of the proposed FDTONN and existing methods carried out in Jupiter notebook in an anaconda environment. The total data is split into 70% training data and 30% test data. The confusion matrix is used to calculate all the parameters such as

precision, recall, false rate, and prediction performance.

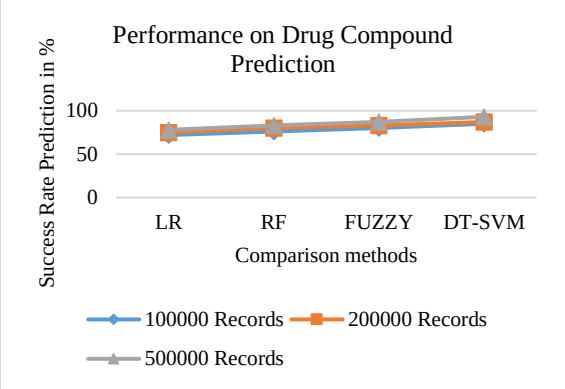

Figure 2: Analysis of Drug success rate prediction

Fig. 2 presents an analysis of drug success rate prediction performance in percentage. The x-axis shows comparison methods, and the y-axis presents prediction performance. The proposed Deep Spectral Neural Classification (FDTONN) algorithm obtained prediction result is 93% for 500000 Records.

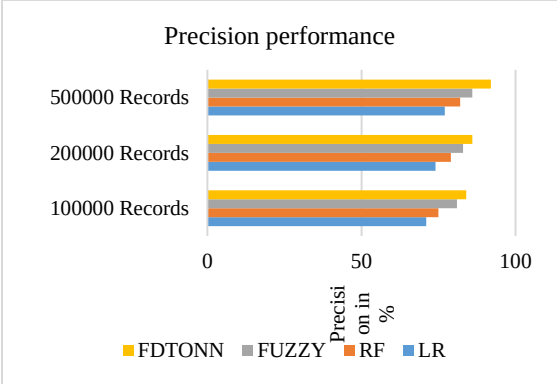

Figure 3: Analysis of Precision performance

Fig. 3 illustrates the analysis of precision performance the recommended and existing results comparison. The x-axis shows the number of data, and the y-axis shows precision performance in percentage.

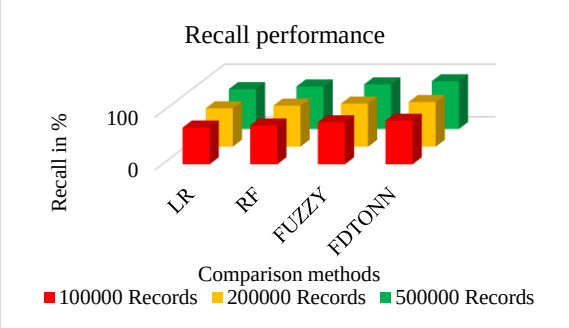

Figure 4: Analysis of Recall performance

Analysis of recall performance the proposed and existing algorithm comparison results are presented in Fig. 4. The proposed FDTONN algorithm recall result has 91%, likewise, the existing algorithm results FUZZY are algorithm has 76%, RF algorithm has 81%, and PrOCTOR has 85%.

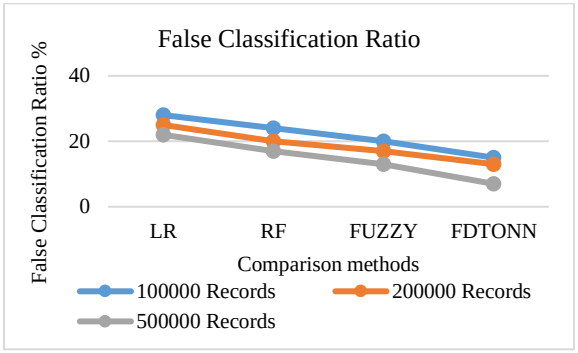

Figure 5: Analysis of false classification ratio performance

Analysis of false rate classification performance results is shown in Fig. 5. The proposed Deep Spectral-FDTONN false rate performance is 7%.

V. CONCLUSION

The proposed method improves the performance of health care monitoring and decision-making. In the future, deep feature-based classification will be used to improve classification accuracy. . Drug feature penetration similarity rate (DFPSR)and deep Fuzzified Decision tree optimized neural network (FDTONN) classifier is applied to find the relational drug combination factor for recommending correct drug to the patients. The selected features are trained with FDTONN. This selected the drug pattern relation based on disease prone weight recommending the best features. The proposed system achieves high performance compared to the other existing system. This prove the proposed system achieved bet result.

REFERENCE

[1] Yuqian Pu, Jiawei Li, Jijun Tang & Fei Guo 2021, 'DeepFusionDTA: drug-target binding affinity prediction with information fusion and hybrid deep-learning ensemble model', IEEE Transaction Computational Biology and Bioinformatics, Pages 1-1, 2021.

[2] Qiwan Hu, Mudong Feng, Luhua Lai1 & Jianfeng Pei1 2018, 'Prediction of drug-likeness

using deep autoencoder neural networks', Frontiers in Genetics', vol.9, pp.1-8.

[3] Song Q, Emergency drug procurement planning based on big-data driven morbidity prediction. IEEE Trans Ind Inf. https ://doi.org/10.1109/tii.2018.28708 79, 2019.

[4] Wong CH, Estimation of clinical trial success rates and related parameters. Bioststics 20(2):273–286, 2019.

[5] Xu X, et. al, Research and development of anticancer agents under the guidance of biomarkers. Cancer Transl Med 5:17–21, 2019.

[6] Zhang C, et al, Assignment optimization of pandemic influenza antiviral drugs in Urban pharmacies, Journal of Ambient Intelligent Humanities Computing, 10:3067–3074. https ://doi.org/10.1007/ s1265 2-018-0872-6, 2019.

[7] Nyabadza F, Coetzee L (2017) A systems dynamic model for drug abuse and drug-related crime in the Western Cape province of South Africa. Comput Math Methods Med 2017:1–13.

[8] Malinzi J 2019, 'Mathematical analysis of a mathematical model of chemovirotherapy: effect of drug infusion method', Computational and Mathematical Methods in Medicine, vol.2019, pp.1–16.

[9] Lee SH, Kang JW, Lin T, Lee JE, Jin DI (2013) Teratogenic potential of antiepileptic drugs in the zebrafish model. BioMed Res Int 2013:1–6.

[10] Zhimiao, Yu, Jiarui, Lu, Yuan, Jin, & Yang, Yang 2021, 'KenDTI: An ensemble model for predicting drug-target interaction by integrating multi-source information', IEEE Transaction on Computational Biology and Bioinformatics, vol. 18, no.4, pp. 1305-1314.

[11] Zhang C, et al, Assignment optimization of pandemic influenza antiviral drugs in Urban pharmacies, Journal of Ambient Intelligent Humanities Computing, 10:3067–3074. https ://doi.org/10.1007/ s1265 2-018-0872-6, 2019.

[12] Hu Q, et. al, Prediction of drug-likeness using deep autoencoder neural networks. Frontiers in Genetics,vol.: 9, page:585,2018.

[13] Eysha saad, et. al, Determining the Efficiency of Drugs Under Special Conditions From Users' Reviews on Healthcare Web Forums, IEEE Access, Pages: 85721-85737,2021.

[14] Lo, AW, Siah, KW, & Wong, CH , 'Machine Learning With Statistical Imputation for Predicting Drug Approvals', Harvard Data Science Review, vol.1, no.1, pp.1-42. 2019

[15] Nicolas Bosc, Francis Atkinson, Eloy Felix, Anna Gaulton, Anne Hersey & Andrew R. Leach 2019, 'Large scale comparison of QSAR and conformal prediction methods and their applications in drug discovery', Journal of Cheminformatics, vol.11, no.4, pp.1-16.